www.ijiarec.com

ISSN:2348-2079

Volume-5 Issue-1

International Journal of Intellectual Advancements and Research in Engineering Computations

Video compression using spiht and dwt

*L.Vigneash, **Anto Jeenia.X, **Harsha Menon, **Kavinilavu.A, **Saranya.D

*Assistant Professor, Department of Electronics and Communication Engineering, Maharaja Engineering College, Avinashi.
Email: lvigneash2504@gmail.com

**Students,
Department of Electronics and Communication Engineering, Maharaja Engineering College, Avinashi.
Email:antojeenia@gmail.com

Abstract-In our work, we present wavelet based video compression algorithms. The motion estimation and compensation, which is an essential part in compression, is based on segment movements. The proposed work based on wavelet transform algorithm like Set Partitioning In Hierarchical Trees (SPIHT) and Discrete Wavelet Transform (DWT). Results of average value of Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE) and comparison chart is obtained using MATLAB. The proposed SPIHT and SWT algorithm achieves very good PSNR values, and MSE which makes the techniques more efficient than the 2D Discrete Cosine Transform (DCT) in the H.264/AVC codec.

Keywords- *Motion estimation - SPIHT - DWT - PSNR -MSE - MATLAB - H.264/AVC codec.*

I. Introduction

The main intention of video coding in most video applications is to reduce the amount of video signal for storing and/or transmission purposes without affecting its visual quality. On root of quality, disks capacity and bandwidth the desired video performance can be achieve. For portable digital video applications, highly integrated real time video compression and decompression solutions are required. In reality motion estimation based encoders are the most usually used in video compression techniques. Such encoders make use of inter frame correlation to provide well organized compression. On the other hand Motion estimation process is computationally expensive. Its real time implementation is tricky and pricey for stored video applications, motion based video coding standard MPEG (Moving Picture Experts Group) was principally urbanized, where the encoding process is typically carried out off line on powerful computers. So it is less suitable to implement as a real time compression technique for a portable recording or communication device (video surveillance camera and fully digital video cameras). In such applications, efficient low cost/complexity implementation is the most noteworthy issue. Thus, researchers turned towards the design of new coders more adapted to new video applications requirements which leads some researchers to look for the exploitation of 3D transforms in order to exploit temporal redundancy.

3D transform coder produces video compression ratio which is close to the motion estimation based coding one with less complex processing. Redundancy has not the same pertinence since the efficiency of 3D transform can reduce as pixel's values variation in spatial or temporal dimensions is not uniform. Often the temporal redundancies are more relevant than spatial one. In order to achieve efficient compression by exploiting more and more the redundancies in the temporal domain. This is the basic purpose of the proposed technique. The proposed technique consists on projecting temporal redundancy of each group of pictures into spatial domain to be combined with spatial redundancy in one representation with high spatial correlation. The obtained representation will be compressed as still image with JPEG coder.

II. Wavelet technique

The discrete wavelet transform (DWT) has gained wide popularity due to its excellent decorrelation property, many modern image and video compression systems embody the DWT as the transform stage. After DWT was introduced, several codec algorithms were proposed to compress the transform coefficients as much as possible. Among them, stationary Wavelet Transform (SWT) and Set

Partitioning in Hierarchical Trees (SPIHT) are the most famous ones.

III. SPIHT

This compression schemes is based on wavelet coding technique. The image is transformed using a discrete wavelet transform. In the beginning, the image is decomposed into four sub bands by cascading horizontal and vertical two channel critically sampled filter banks. This process of decomposition continues until some final scale is reached. In each scale there are three sub bands and one lowest frequency sub band. Then Successive Approximation quantization (SAQ) is used to perform embedding coding. This particular configuration is also called QMF pyramid. The SPIHT algorithm is used to the multi resolution pyramid after the subband/wavelet transformation is performed.

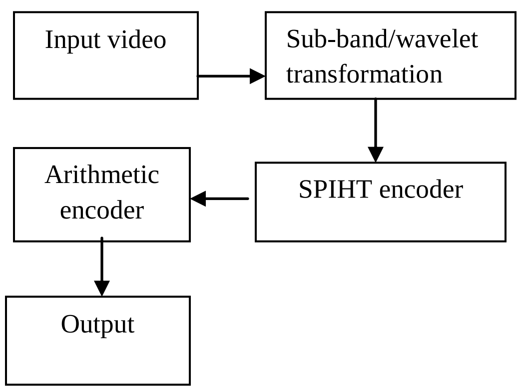

Fig 3.1 SPHIT Video Coding System

IV. Discrete Wavelet Transform (DWT)

The wavelet transform of a signal gives its time frequency representation. On the wavelet transform which was developed to overcome the shortcomings of the Short Time Fourier Transform (STFT) is a multiresolution transformation. This property of the Fourier transform helps to analyse different frequencies of signals at different resolutions. In contrast to a wave which is an oscillating function of time or space the wavelets are localized waves and are non periodic. The energy of the wavelets is concentrated in time or space.

DWT (Discrete Wavelet transformation) is one of a most famous transformation method, here we try to indentify the moving/changing and non-moving pixels between frames and try to eliminate the stagnant pixels [5] and thereby reducing the video size. DWT method can be found efficient or more useful for compressing high definition slow motion videos as the motion between frame to frame is very minimal and also high definition slow motion videos has higher bit rate compared to the fast motion videos [6]. Usage of DWT method is not much recommended or seen beneficial in case of Wyner-Ziv [7] where the focus is more on low end or low cost videos. In most of the Wyner-Ziv based transcoder a low complexity MPEG-2 to H.264 [8]. Usage of Discrete wavelet Transformation (DWT) can also be said more beneficial when compared to Discrete cosine transformation (DCT) where the focus is more on the content modeling[9, 10] due to the known drawback that surface with the usage of DCT.

For Compression methods involving discrete wavelet transformation two types of redundancies are to be considered. They are
(a) Spatial redundancy and
(b) Temporarily redundancy.
Spatial redundancy is also called as inter-frame redundancy and it is said when there is a redundancy indentified between pixel (i.e., found mostly on gray scale images or black and white videos). Temporarily redundancy is also called as intra- frame redundancy and it is said when there is a redundancy identified between frames to frame. (i.e., found mostly on slow motion videos where the moving pixels are less over the frames). Any video is a combination of the set of frames which is a set of digital images that are timed to run approx 55 digital images per second and 55x60 frames that runs for a minute and forms 1 minute of video. When any two consecutive frames in a video are seen the difference between the two would be very minimal/negligible and it could not be seen/identified for human eye. Video transcoding plays a crucial role in identifying such changes and convert the video from one format to another format occupying low disk space. Figure-1 shown below is the end to end flow

A raw video or the uncompressed original video has to pass through multiple layers of internal processing in the coding before it comes out as compressed video with less disk space. Once the videos are compressed they are easily transported/shared over the internet and also they can be easily uploaded or downloaded. The input source video is loaded to coding which is then passed to discrete wavelet transformation (DWT) based transcoder/compression process and then the decompressed readable form of output video with less size is produced.

With the use of DWT process high compression, high resolution and efficiency of the video is increased, same is also shown with the calculation of PSNR value which is much improved compared to DCT based transponders.

Frame separation is a process in which the input video is separated in to individual frames (i.e., approx 1 second of video is separated in to 55 frames). Each frame looks like a digital still image. At this stage the frames are separated as i-frames, p-frames and b-frames. I-frame is also known as intra frames and these are most common frames that are independently decoded to any other frames. P-frames are known as predictive frames in forward direction and b-frames are known as backward predictive or bi-directional frames. Once the frames are separated that are loaded in to coding and kept on buffer for further processing to eliminate the redundant pixels.

This is typically called as foreground detection technique. This is the key component of the compression mechanism where it essentially identifies the redundant pixels. This technique is based on the static pixels which do not get to change between frames to frame. A motion is detected only when the foreground is moving and segmented from the background. Once the non-moving units are successfully separated to those that have the motion next process of motion compensation is applied.

Coding units at this stage have already passed through frame separation and background separation process. In motion compensation process only those frames and those macro blocks that are identified with the motion are taken in buffer. Initially before processing for motion compensation, motion estimation is carried which would give a tentative estimation on how many frames or macro blocks that can be compensated using block matching algorithm. In motion estimation the movements of macro blocks are identified by using p-frame (predictive frames). Those frames that have minimal movement of macro blocks are compensated with the p-frame and b-frame. Motion compensation is only possible on p- frames where the motion could be predicted, however it cannot be applied on I-frame which is a completed independent frame.

Once the motion compensation process is completed the video now need to be built from the buffers that are constructed out from background separation and motion compensation process. All the redundant frames are eliminated at this stage and resultant video is in compressed form lesser in size compared to the original video. Efficiency of the compression ratio and bit stream is calculated with the new PSNR value. The resultant video is not necessary to be of separate format, it can be of the same format of the input video also.

V. Set Partitioning In Hierarchcal Trees (SPIHT)

The SPIHT algorithm is a more efficient implementation of EZW (Embedded Zero Wavelet) algorithm which was presented by Shapiro. After applying wavelet transform to an image, the SPIHT algorithm partitions the decomposed wavelet into significant and insignificant partitions based on the following function,

$$S_n(T) = \begin{cases} 1, max_{(i,j)\epsilon T}\{|c_{i,j}|\} \geq 2^n \\ 0, Otherwise \end{cases}$$

Here Sn(T) is the significance of a set of coordinates T, and ci,j is the coefficient value at coordinate (i, j). There are two passes in the algorithm- the sorting pass and the refinement pass. The SPIHT encoding process utilizes three lists LIP (List of Insignificant Pixels) – It contains individual coefficients that have magnitudes smaller than the thresholds LIS (List of Insignificant Sets) – It contains set of wavelet coefficients that are defined by tree structures and are found to have magnitudes smaller than the threshold.
LSP (List of Significant Pixels) – It is a list of pixels found to have magnitudes larger than the threshold (significant).
The sorting pass is performed on the above three lists. The maximum number of bits required to represent the largest coefficient in the spatial orientation tree is obtained and represented by nmax which is ,

$$n_{max} = \lfloor log_2(max_{i,j}\{|c_{i,j}|\})\rfloor$$

During the sorting pass, those coordinates of the pixels which remain in the LIP are tested for significance by using equation 2. The result is sent to the output and out of it the significant will be transferred to the LSP as well as having their sign bit output. Sets in the LIS will get their significance tested too and if found significant, will be removed and partitioned into subsets. Subsets with only one coefficient and found to be significant, will be eliminated and divided into subsets. Subsets having only one coefficient and found to be significant will be inserted to the LSP, otherwise they will be inserted to the LIP. In the refinement pass, the nth MSB of the coefficients in the LSP is the final output. The value of n is decremented and the sorting and refinement passes are applied again. These passes will keep on continuing until either the desired rate is reached or n =0, and all nodes in the LSP have all their bits output. The latter case will give an almost exact reconstruction since all the coefficients have been processed completely. The bit rate can be controlled exactly in the SPIHT algorithm as the output produced is in single bits and the algorithm can be finished at any time. The decoding process follows the encoding exactly and is almost symmetrical in terms of processing time.

One of the most efficient algorithms in the area of image compression is the Set Partitioning in Hierarchical Trees (SPIHT). In essence it uses a sub-band coder, to produce a pyramid structure where an image is decomposed sequentially by applying power complementary low pass and high pass filters and then decimating the resulting images. These are one dimensional filters that are applied in cascade (row then column) to an image whereby creating a four way decomposition LL (low pass then another low pass), LH (low pass then high pass), HL (high and low pass) and finally HH (high pass then another high pass). The resulting LL version is again four-way decomposed.

There exists a spatial relationship among the 1-3,6 coefficients at different levels and frequency sub bands in the pyramid structure. A wavelet coefficient at location (i,j) in the pyramid representation has four direct descendants (offsprings) at locations:

$$O(i,j)=\{(2i,2j),(2i,2j+1),(2i+1,2j),(2i+1,2j+1)\}$$

and each of them recursively maintains a spatial similarity to its corresponding four offspring. This pyramid structure is commonly known as spatial orientation tree. The similarity among sub-bands within levels in the wavelet space 6 . If a given coefficient at location (i,j) is significant in magnitude then some of its descendants will also probably be significant in magnitude. The SPIHT algorithm takes advantage of the spatial similarity present in the wavelet space to optimally find the location of the wavelet coefficient that are significant by means of a binary search algorithm.

The SPIHT algorithm sends the top coefficients in the pyramid structure using a progressive transmission scheme. This scheme is a method that allows obtaining a high quality version of the original image from the minimal amount of transmitted data. The pyramid wavelet coefficients are ordered by magnitude and then the most significant bits are transmitted first, followed by the next bit plane and so on until the lowest bit plane is reached. It has been shown that progressive transmission can significantly reduced the Mean Square Error (MSE) distortion for every bitplane sent To take advantage of the spatial relationship among the coefficients at different levels and frequency bands, the SPIHT coder algorithm orders the wavelets coefficient according to the significance test defined as

$$\max_{(i,j)\in\tau_m} \left|C_{i,j}\right| \geq 2^n$$

where C is the wavelet coefficient at the nth bit plane, at location (i,j)of the i,j τ m subset of pixels, representing a parent node and its descendants. If the result of the significance test is yes an S flag is set to 1 indicating that a particular test is significant. If the answer is no, then the S flag is set to 0, indicating that the particular coefficient is insignificant.

$$S_n(\tau)=\begin{cases}1, & \max_{(i,j)\in\tau}\left|C_{i,j}\right|\geq 2^n\\ 0, & \text{otherwise}\end{cases}$$

Wavelets coefficients which are not significant at the nth bitplane level may be significant at (n-1)th bit plane or lower. This information is arranged, according to its significance, in three separate lists: list of insignificant sets (LIS), the list of insignificant.

Pixels (LIP) and the list of significant pixels (LSP). In the decoder, the SPIHT algorithm replicates the same number of lists. It uses the basic principle that if the execution path of any algorithm is defined by the results on its branching points and if the encoder and decoder have the same sorting algorithm then the decoder can recover the ordering information easily.

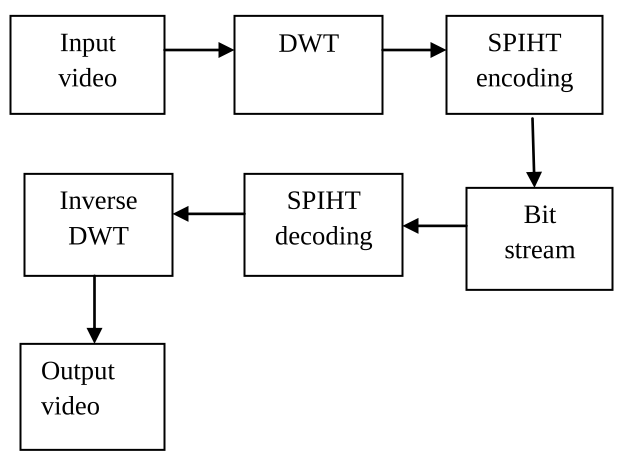

Fig 5.1 Block Diagram of SPIHT Video Compression

Motion estimation is the process of determining motion vectors that describe the transformation from one 2D image to another;

usually from adjacent frames in a video sequence. It is an ill-posed problem as the motion is in three dimensions but the images are a projection of the 3D scene onto a 2D plane. The motion vectors may relate to the whole image (global motion estimation) or specific parts, such as rectangular blocks, arbitrary shaped patches or even per pixel . The motion vectors may be represented by a translational model or many other models that can approximate the motion of a real video camera, such as rotation and translation in all three dimensions and zoom.

Motion compensation is an algorithmic technique used to predict a frame in a video, given the previous and/or future frames by accounting for motion of the camera and/or objects in the video. It is employed in the encoding of video data for video compression , for example in the generation of MPEG-2 files. Motion compensation describes a picture in terms of the transformation of a reference picture to the current picture. The reference picture may be previous in time or even from the future. When images can be accurately synthesised from previously transmitted/stored images, the compression efficiency can be improved.

VI.Result Analysis

In simulation we have consider standard video sequences (Foreman.YUV, Akio.YUV etc). The videos we consider has few scene change, so we consider every fifth frame as a key frame for the coding. To check the compression performance we use the compression gain formula as given bellow,

$$CR = \frac{(NxN)xF_n xR_t}{[(NxN)xKF_n xR_{t1}]+[(NxN)xDF_n xR_{t2}]}$$

CR - Compression ration.
N - Frame size.
F_n - Total number of frames.
KF_n - Total number of key frames.
DF_n - Total number of delta frames.
R_t - Number of bits per pixel in original video.
R_{t1} - Number of bits per pixel assign for key frame in compressed video.
R_{t2} - Number of bits per pixel assign for delta frame in compressed video.

It is to be noted that in our simulation we use

$$R_{t2} = 0.3R_{t1}$$

Bit Per Pixel for Key Frame = 0.9, Bit Per Pixel

Parameters	Size (bytes)	Total bit rate(kbps)	Frame rate(frames/sec)
Input video	6,917,120	6912	15
Compressed video	2,947,464	4715	10

Table 1. Compression Ratio

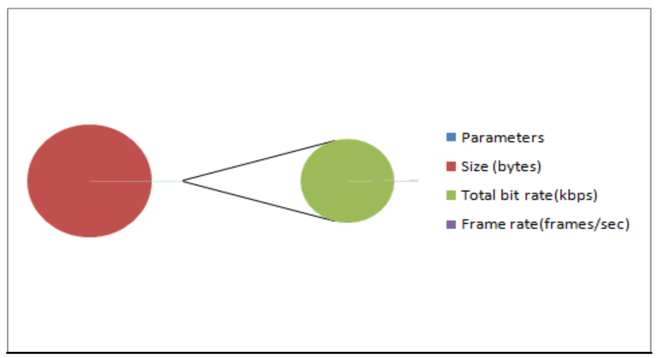

Fig 1 SPIHT Compression

VII.Conclusion

Proposing a simple and effective method combined with Huffman encoding for further compression in this paper that saves a lot of bits in the image data transmission. There are very wide range of practical value for today that have large number of image data's to be transmitted. We have implemented the SPIHT algorithm using MATLAB and its wavelet toolbox. Student's familiarity with this platform is what guided us through this development. Further research includes the optimization of the code as in 8, improvements for region of interest coding, and real time implementation of the algorithm. Although the program runs slower than its C counterpart, it gives the flexibility of using stable and robust set of tools such as MATLAB.

Reference

1) Thazni Aziz, D. Raveena Judie Dolly “Motion Estimation and Motion Compensated Video Compression Using DCT And DWT”, International Journal of Emerging Technology and Advanced Engineering,(ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 2, Issue 12, December 2012).
2) Kohtaro Asai, Tokumichi Murakami, Shuichi Yamagishi, Akira Minezawa, Yusuke Itani, “New Video Coding Scheme Optimized for High-Resolution Video Sources”, IEEE journal of selected topics in signal processing, Vol. 5, No. 7, November 2011.
3) Dishant Khosla, Amandeep Kaur, ”Design of Hybrid Compression Model using DWT-DCTHUFFMAN Algorithms for Compression of Bit Stream”,International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 5, July – 2012.
4) MuzhirShaban, Al-Ani1andTalal Ali Hammour, “Video Compression Algorithm Based on Frame Difference Approaches”, International Journal on Soft Computing (IJSC) Vol.2, No.4, November 2011.
5) Limin Liu, Member, Zhen Li, and Edward J. Delp, “ Efficient and Low-Complexity Surveillance Video Compression Using Backward Channel Aware Wyner Ziv Video Coding, IEEE transactions on circuits and systems for video technology, vol. 19, no. 4, April 2009.
6) C.-L. Chang and B. Girod, “Direction-Adaptive Discrete Wavelet Transform for image compression,” Image Processing, IEEE Transactions on, vol. 16, no. 5, pp. 1289–1302, May 2007